

Corpus Query Language (CQL): Quick Reference Card

KonText interface – first steps

- selection of a corpus
- query type
 - simple: just a word or phrase
(with match case & RE on/off switches)
 - advanced: full CQL
- search restriction (optional)

Corpus as viewed by KonText

- corpus is a vertical sequence of *tokens*
- tokens have a set of *positional attributes**
- sequence of tokens segmented by *structures**
- structures have a set of *structural attributes**

* corpus-dependent

Example

```
<sp sex="MALE" name="Bassanio">
...
<s id="s-111">
I    PRON    Case=Nom    I    nsubj    +3
will AUX     VerbForm=Fin will aux     +2
not  PART    _          not  advmod   +1
fail VERB    VerbForm=Inf fail parataxis -4
you  PRON    Case=Acc   you  obj      -1
.    PUNCT   _          .    punct   -6
</s>
...
</sp>
```

Regular expressions: standard characters

(examples)

```
a ..... letter a
5 ..... number 5
, ..... comma
```

Regular expressions: special characters

```
. ..... any single character
[ ] ..... any single char from those in brackets
[^ ] .. any single char except for those in brackets
| ..... alternative
( ) ..... grouping
\ ..... treat the following char as a regular one
```

Regular expressions: quantifiers

```
* ..... zero or more repetitions of a previous char
+ ..... one or more repetitions of a previous char
? ..... zero or one repetition of a previous char
{n} ..... exactly n repetitions of a previous char
{n,} .... n or more repetitions of a previous char
{n, k} . from n to k repetitions of a previous char
```

Examples

```
dog ..... the word dog
19.. ..... any four-char string starting with 19
re.* ..... any string starting with re
re.*on ..... any string starting re and ending on
(ha){2,4} ..... haha, hahaha or hahahaha
[aeiouy]+ ..... any combination of vocals
(a|e|i|o|u|y)+ ..... equivalent of the above
```

CQL: within a single token

(examples)

```
[word="mean"]
... all instances of the word form mean
[word="mean" & upos="VERB"]
... all verbal instances of mean
[word="mean" & upos!="VERB"]
... all non-verbal instances of mean
[word="mean" | word="average"]
... words mean or average
[(word="mean" | word="average") & upos="ADJ"]
... all adjectival instances of mean / average
```

CQL: beyond a single token

(examples)

```
[lemma="big"] [lemma="dog"]
... matches big dog
[lemma="big"] [] [lemma="dog"]
... big dog with an intervening token
[lemma="big"] []{0,5} [lemma="dog"]
... big dog with up to 5 intervening tokens
[lemma="big"] []{0,5} [lemma="dog"] within <s/>
... ditto, but within a single sentence
<s> [upos="INTJ"] within <sp name="Juliet"/>
... interjection starting a sentence in Juliet's speech
1:[upos="ADJ"] 2:[upos="ADJ"] & 1.word=2.word
... two identical adjectives following each other
```